УДК 621.391

С. Д. Штовба, д-р техн. наук, проф.;
О. Д. Панкевич, канд. техн. наук, доц.;
А. В. Нагорна, асп.

КРИТЕРІЇ НАВЧАННЯ НЕЧІТКОГО КЛАСИФІКАТОРА, ЩО ВРАХОВУЮТЬ ПЛАТІЖНУ МАТРИЦЮ

В нечітких класифікаторах зв'язок «входи—вихід» описується лінгвістичними правилами «Якщо — тоді», антецеденти яких містять нечіткі терми «низький», «середній», «високий» тощо. Для підвищення безпомилковості нечіткий класифікатор навчають за експериментальними даними. Узагальнено критерії навчання нечіткого класифікатора на випадок платіжної матриці, в якій записані вартості помилок різних типів. Комп'ютерні експерименти із розв'язання задачі діагностики хвороби серця показали, що найкращу якість настроювання забезпечує використання критерію навчання, в якому відстань між нечіткими результатами логічного виведення та експериментальними даними для випадків помилкової класифікації зважується штрафним коефіцієнтом.

Вступ

Задача класифікації полягає у віднесенні об'єкта за деякими ознаками до одного з класів. Останнім часом все популярнішими стають класифікатори на основі нечіткої бази знань. В нечітких класифікаторах зв'язок «входи—вихід» описується лінгвістичними правилами <Якщо—тоді>, антецеденти яких містять нечіткі терми «низький», «середній», «високий» тощо.

Для підвищення безпомилковості нечіткий класифікатор навчають за експериментальними даними. Для цього ітераційно змінюють його параметри, щоб мінімізувати відстань між експериментальними даними та результатами нечіткого виведення. Цю відстань, яку назвемо критерієм навчання, можна визначити у різний спосіб. В статті [1] порівнюються 3 критерії навчання нечіткого класифікатора на основі частоти помилок, відстані між нечіткими результати класифікації і експериментальними даними та їх комбінації. Наведені в [1] критерії навчання відповідають випадку однакової вартості помилок класифікації різних типів. На практиці часто зустрічаються задачі, коли вартості помилок різних типів суттєво різняться. Наприклад, вартість помилок першого роду (пропуск цілі) в кілька разів більша за вартість помилок другого роду (хибна тривога).

Для врахування помилок різних типів під час навчання класифікаторів використовуються 2 підходи. За першим підходом в постановку задачі навчання вводять обмеження на допустимі рівні помилок найнебезпечніших типів. У випадку лише двох типів помилок у відповідності до леми Неймана—Пірсона мінімізують кількість помилок другого роду за обмеження на рівень помилок першого роду [2]. За другим підходом мінімізують сумарні збитки від помилкової класифікації. В цьому випадку вважають відомими вартості помилок кожного типу, які зазвичай задають платіжною матрицею.

Метою статті є узагальнення критеріїв навчання нечіткого класифікатора з статті [1] з урахуванням платіжної матриці. Крім того, за результатами комп'ютерних експериментів визначається, який з цих критеріїв забезпечує найкращу якість навчання нечіткого класифікатора.

1. Нечіткий класифікатор

Позначимо через $\mathbf{X}=(x_1, x_2,\ldots,x_n)$ — вектор інформативних ознак об'єкта класифікації, а через $l_1, l_2, ..., l_m$ — класи рішень. Тоді задачі класифікації відповідатиме таке відображення: $\mathbf{X}=(x_1, x_2,\ldots,x_n) \to y\in\{l_1, l_2, ..., l_m\}$, де y — результат класифікації. Ґрунтуючись на [3] нечітку базу знань цього відображення запишемо так:

$$\text{Якщо}\left(x_1 = \tilde{a}_{1j} \text{ та } x_2 = \tilde{a}_{2j} \text{ та ... та } x_n = \tilde{a}_{nj} \text{ з вагою } w_j\right), \text{тоді } y = d_j, \quad j = \overline{1, k}, \tag{1}$$

де k — кількість правил; $d_j \in \{l_1, l_2, ..., l_m\}$ — значення консеквента j-го правила; $w_j \in [0, 1]$ — ваговий коефіцієнт, який задає достовірність j-го правила, $j = \overline{1, m}$; $\tilde{a}_{ij}$ — нечіткий терм, яким оцінюється ознака x_i в j-му правилі $i = \overline{1, n}$, $j = \overline{1, k}$.

Класифікація поточного об'єкта, який задано вектором $\mathbf{X}^* = \left(x_1^*, x_2^*, \ldots, x_n^*\right)$, здійснюється таким чином. Спочатку розраховується ступінь виконання j-го правила з бази знань (1):

$$\mu_j\left(X^*\right) = w_j \cdot \left(\mu_j\left(x_1^*\right) \wedge \mu_j\left(x_2^*\right) \wedge ... \wedge \mu_j\left(x_n^*\right)\right), \quad j = \overline{1, k}, \tag{2}$$

де $\mu_j\left(x_i^*\right)$ — ступінь належності значення x_i^* нечіткому терму $\tilde{a}_{ij}$; $\wedge$ — t-норма, яку зазвичай реалізують операцією мінімуму або добутком.

Ступінь належності вхідного вектора $\mathbf{X}^*$ до класів $l_1, l_2, ..., l_m$ розраховується так:

$$\mu_{l_s}\left(y^*\right) = \underset{\forall j:\, d_j = l_s}{\text{agg}} \left(\mu_j\left(X^*\right)\right), \quad s = \overline{1, m}, \tag{3}$$

де agg — агрегування нечітких висновків за окремими правилами бази знань.

Агрегування реалізуємо операцією максимуму над ступенями належності, що відповідає схемі логічного виведення з єдиним правилом переможцем (single winner rule) [4].

Нечітким рішенням задачі класифікації буде нечітка множина

$$\tilde{y}^* = \left(\frac{\mu_{l_1}\left(y^*\right)}{l_1}, \frac{\mu_{l_2}\left(y^*\right)}{l_2}, ..., \frac{\mu_{l_m}\left(y^*\right)}{l_m}\right). \tag{4}$$

Результатом логічного виведення оберемо ядро нечіткої множини (4), тобто клас з максимальним ступенем належності:

$$y^* = \underset{\{l_1, l_2, ..., l_m\}}{\arg} \max_{s = \overline{1, m}} \left(\mu_{l_s}\left(y^*\right)\right).$$

Можлива ситуація, коли в ядро нечіткої множини (4) входять кілька елементів. Тоді об'єкт одночасно належить кільком класам з однаковими ступенями, значення яких дорівнює $\max\limits_{s = \overline{1, m}} \left(\mu_{l_s}\left(y^*\right)\right)$. Для вибору одного з цих конкурентних класів застосуємо схему голосування правил [4]. За нею для кожного із конкурентних класів розрахуємо суму ступенів (2) виконання відповідних правил. Рішенням обираємо клас з максимальною сумою.

2. Критерії навчання без урахування платіжної матриці

Навчальну вибірку з M пар «входи—вихід» запишемо так:

$$\left(\mathbf{X}_r, y_r\right), \quad r = \overline{1, M}, \tag{5}$$

де $y_r \in \{l_1, l_2, ..., l_m\}$.

Введемо такі позначення:

$\mathbf{P}$ — вектор параметрів функцій належності термів з бази знань (1); $\mathbf{W}$ — вектор вагових коефіцієнтів правил бази знань (1); $F\left(\mathbf{K}, \mathbf{X}_r\right) \in \{l_1, l_2, ..., l_m\}$ — результат класифікації за базою знань з параметрами $\mathbf{K} = (\mathbf{P}, \mathbf{W})$ для вхідного вектора $\mathbf{X}_r$ з r-го рядка вибірки (5).

Навчання нечіткого класифікатора полягає в знаходженні вектора $\mathbf{K}$, який мінімізує відстань між результатами логічного виведення та експериментальними даними з вибірки (5). В [1] розглянуто 3 способи задання цієї відстані в формі критеріїв навчання нечіткого класифікатора.

Критерій 1. За відстань між бажаною та дійсною поведінкою моделі обрано частоту помилок класифікації:

$$Crit_1 = \frac{1}{M} \sum_{r = \overline{1, M}} \Delta_r(\mathbf{K}), \tag{6}$$

де $\Delta_r(\mathbf{K}) = \begin{cases} 1, & \text{якщо} \;\; y_r \neq F(\mathbf{K}, \mathbf{X}_r); \\ 0, & \text{якщо} \;\; y_r = F(\mathbf{K}, \mathbf{X}_r). \end{cases}$

Переваги критерію (6) полягають в його простоті та ясній змістовній інтерпретації. Частота помилок застосовується як критерій навчання різноманітних систем розпізнавання образів. Але цільова функція в задачі оптимізації за цим критерієм приймає дискретні значення, що ускладнює застосування швидких градієнтних методів оптимізації, особливо за малих вибірок даних.

Критерій 2. В роботі [5] запропоновано для розрахунку відстані між експериментальними даними та результатами нечіткого моделювання значення вихідної змінної y в навчальній вибірці фаззіфікувати таким чином:

$$\left.\begin{array}{ll} \tilde{y} = \left(\dfrac{1}{l_1}, \dfrac{0}{l_2}, \ldots, \dfrac{0}{l_m}\right), & \text{якщо} \;\; y = l_1; \\ \tilde{y} = \left(\dfrac{0}{l_1}, \dfrac{1}{l_2}, \ldots, \dfrac{0}{l_m}\right), & \text{якщо} \;\; y = l_2; \\ \ldots\ldots\ldots\ldots\ldots\ldots & \ldots\ldots\ldots\ldots \\ \tilde{y} = \left(\dfrac{0}{l_1}, \dfrac{0}{l_2}, \ldots, \dfrac{1}{l_m}\right), & \text{якщо} \;\; y = l_m. \end{array}\right\} \tag{7}$$

Критерій навчання базується на відстані між результатами логічного виведення у вигляді нечіткої множини (4) та бажаними нечіткими значеннями вихідної змінної (7):

$$Crit_2 = \sqrt{\frac{1}{M} \sum_{r=\overline{1, M}} D_r(\mathbf{K})}\,, \tag{8}$$

де $D_r(\mathbf{K}) = \sum_{s=\overline{1, m}} \left(\mu_{l_s}(y_r) - \mu_{l_s}(\mathbf{K}, \mathbf{X}_r)\right)^2$ — відстань між бажаною та дійсною вихідними нечіткими множинами при класифікації r-го об'єкта з навчальної вибірки (5); $\mu_{l_s}(y_r)$ — ступінь належності значення змінної y з r-го рядка навчальної вибірки до класу l_s згідно до (5); $\mu_{l_s}(\mathbf{K}, \mathbf{X}_r)$ — розрахований за формулою (3) ступінь належності виходу нечіткої моделі з параметрами $\mathbf{K}$ до класу l_s за вхідного вектора $\mathbf{X}_r$.

Цільова функція в задачі навчання за критерієм (8) не має довгих плато, тому вона придатна до оптимізації градієнтними методами. Але в деяких випадках [1, 3, 6] оптимальна за (8) нечітка база знань, не забезпечує близький до мінімального відсоток помилок класифікації (6). Це пояснюється тим, що близькі до границь класів об'єкти вносять майже однаковий вклад в критерій навчання (8) як за правильної, так і за помилкової класифікації.

Критерій 3. Цей критерій успадковує переваги двох попередніх. Ідея полягає в збільшенні відстані D для помилково класифікованих об'єктів:

$$Crit_3 = \sqrt{\frac{1}{M} \sum_{\text{r}=\overline{1, M}} \left(\Delta_r(\mathbf{K}) \cdot penalty + 1\right) \cdot D_r(\mathbf{K})}\,, \tag{9}$$

де $penalty > 0$ — штрафний коефіцієнт.

Під час навчання за критерієм (9) вибір напрямку крокування до оптимуму найбільшою мірою залежить від помилково класифікованих об'єктів. Така поведінка схожа на адаптивний метод оптимізації з [7], коли для повторного навчання частіше пред'являють помилково розпізнанні об'єкти.

3. Критерії навчання із врахуванням платіжної матриці

Платіжною називається така квадратна матриця:

$$\mathbf{C}=\begin{bmatrix} 0 & c(l_1,l_2) & ... & c(l_1,l_m) \\ c(l_2,l_1) & 0 & ... & c(l_2,l_m) \\ \dots\dots\dots & \dots\dots\dots & \dots & \dots\dots\dots \\ c(l_m,l_1) & c(l_m,l_2) & ... & 0 \end{bmatrix}, \tag{10}$$

де $c(l_i, l_j)$ — ціна помилки типу $l_i \to l_j$, коли замість правильного рішення l_i під час класифікації помилково обирається рішення l_j, $i=\overline{1,m}$, $j=\overline{1,m}$, $i \neq j$. Нулі на головній діагоналі матриці (10) вказують на відсутність плати за правильну класифікацію.
За відомої платіжної матриці (10) критерій 1 перетворюється на такий:

$$Crit_{1C}=\frac{1}{M}\sum_{r=\overline{1,M}} c(u,v) \to \min, \tag{11}$$

де $u=y_r$ та $v=F(\mathbf{K},\mathbf{X}_r)$.

Критерій 2 модифікуємо таким чином, щоб для розрахунку відстані $D_r(\mathbf{K})$ координати зважити вартостями відповідних помилок. В результаті формула (8) перетвориться на таку:

$$Crit_{2C}=\sqrt{\frac{1}{M}\sum_{r=\overline{1,M}} D_r(\mathbf{K},\mathbf{C})}, \tag{12}$$

де $D_r(\mathbf{K},\mathbf{C})=\sum\limits_{s=\overline{1,m}}\left(1+c\left(y_r,l_s\right)\right)\cdot\left(\mu_{l_s}\left(y_r\right)-\mu_{l_s}\left(\mathbf{K},\mathbf{X}_r\right)\right)^2$ — зважена відстань між бажаною та дійсною вихідними нечіткими множинами за класифікації r-го об'єкта з навчальної вибірки (5).

В критерій 3 замість $D_r(\mathbf{K})$ введемо зважену відстань $D_r(\mathbf{K},\mathbf{C})$ з (12):

$$Crit_{3C}=\sqrt{\frac{1}{M}\sum_{\mathrm{r}=\overline{1,M}}\left(\Delta_r(\mathbf{K})\cdot penalty+1\right)\cdot D_r(\mathbf{K},\mathbf{C})}.$$

4. Комп'ютерні експерименти

Розглядається задача діагностики хвороби серця за даними з Statlog Heart Data Set [8]. В кожному із 270 рядків цієї бази даних описано 13 ознак стану пацієнта, за якими приймають рішення про наявність або відсутність хвороби серця. Відома така платіжна матриця: $\mathbf{C}=\begin{bmatrix} 0 & 1 \\ 5 & 0 \end{bmatrix}$. Вона вказує, що вартість пропуску цілі в 5 разів вища, ніж хибної тривоги.

Навчальну вибірку сформуємо з рядків бази даних, що містять граничні значеннями кожного із 13-ти атрибутів. Додатково в навчальну вибірку включимо усі непарні рядки бази даних. Решту даних занесемо у тестову вибірку. В результаті отримуємо навчальну вибірку з 145 рядків і тестову — з 125.

Спроектуємо нечіткий класифікатор з трьома входами:

x_1 — вік; x_{10} — старечий пік, обумовлений реакцією на фізичне навантаження (old peak); x_{12} — кількість крупних судин, зафарбованих за результатами рентгеноскопії.

За розподілом даних сформовано нечітку базу знань (табл. 1). Нечіткі терми задано гаусовою функцією належності:

$$\mu(x)=\exp\left(-\frac{(x-b)^2}{2c^2}\right),$$

де b — координата максимуму та $c>0$ — коефіцієнт концентрації.

Параметри функцій належностей початкового нечіткого класифікатора наведено в табл. 2.

Таблиця 1

Нечітка база знань

№	x_1	x_{10}	x_{12}	y
1	—	Низький	Мало	Здоровий
2	Молодий	Високий	Мало	Здоровий
3	—	—	Багато	Хворий
4	Старий	Високий	Мало	Хворий

Таблиця 2

Параметри функцій належності нечітких класифікаторів

Змінна	Терм	Початковий		Best_min		Best_prod	
		b	c	b	c	b	c
x_1	Низький	29	20,4	29	9,5	29	8,36
	Високий	77	20,4	77	9,83	77	8,43
x_{10}	Низький	0	2,63	0	0,6	0	2,57
	Високий	6,2	2,63	6,2	2,93	6,2	0,86
x_{12}	Низький	0	1,27	0	1	0	1,34
	Високий	3	1,27	3	2	3	1,2

Для кожного критерію проведемо 500 експериментів із навчання нечіткої бази знань на основі квазіньютонівського алгоритму. Після навчання кожний класифікатор перевіримо на тестовій вибірці за критерієм (11). Експерименти проведемо для двох нечітких баз знань, в одній із яких *t*-норма реалізована операцією мінімуму, а в іншій — добутком.

Під час навчання налаштуємо вагові коефіцієнти кожного із 4-х правил бази знань та коефіцієнти концентрації (c) функції належності кожного нечіткого терма. В базі знань усі нечіткі терми є крайніми, тому згідно з [9] координати максимумів функцій належності — параметри b — під час навчання не будемо налаштовувати і залишимо їх рівними границям діапазону зміни вхідних змінних. Таким чином, загальна кількість настроюваних параметрів становить $4+6=10$. Початкові точки для навчання обиралися випадково — для вагових коефіцієнтів правил з діапазону $[0,1]$, а для параметрів функцій належності в межах $\pm 20\ \%$ від значень з табл. 2.

Після 3000 експериментів ми отримали множину з 8 класифікаторів, кожен з яких забезпечує мінімальне значення критерію (11) на тестовій вибірці на рівні 0,424. З цієї множини ми обрали 2 класифікатори Best_min та Best_prod, які мають мінімальне значення цього критерію і на навчальній вибірці, а саме $Crit_{1C}(\text{Best\_min})=0,566$ та $Crit_{1C}(\text{Best\_prod})=0,6$. Тут і надалі через Best_min позначено найкращий класифікатор, у якого *t*-норма реалізовано операцією мінімуму, а через Best_prod — найкращий класифікатор, у якого *t*-норма реалізовано добутком. Вагові коефіцієнти правил у Best_min дорівнюють $\omega_1=1$, $\omega_2=0{,}45$, $\omega_3=1$ та $\omega_4=0{,}83$, а у Best_prod — $\omega_1=0{,}39$, $\omega_2=0{,}55$, $\omega_3=0{,}8$ та $\omega_4=1$. Параметри функцій належності цих класифікаторів зведені в табл. 2.

Результати експериментів (табл. 3 та рис. 1, 2) свідчать, що в середньому краща якість навчання спостерігається у разі використання критерію $Crit_{3C}$. Розкид результатів навчання найширший за оптимізації за критерієм $Crit_{1C}$ і найвужчій у випадку критерію $Crit_{2C}$.

Таблиця 3

Статистика навчання нечітких класифікаторів

t-норма	Критерій навчання	Значення критерію $Crit_{1C}$ на тестовій вибірці			
		мінімальне	середнє	максимальне	СКВ
мінімум	$Crit_{1C}$	0,424	1,185	1,408	0,309
	$Crit_{2C}$	0,96	1,081	1,112	0,015
	$Crit_{3C}$	0,424	0,614	1,248	0,105
добуток	$Crit_{1C}$	0,424	1,051	1,408	0,399
	$Crit_{2C}$	0,832	0,923	0,952	0,019
	$Crit_{3C}$	0,424	0,627	0,928	0,11

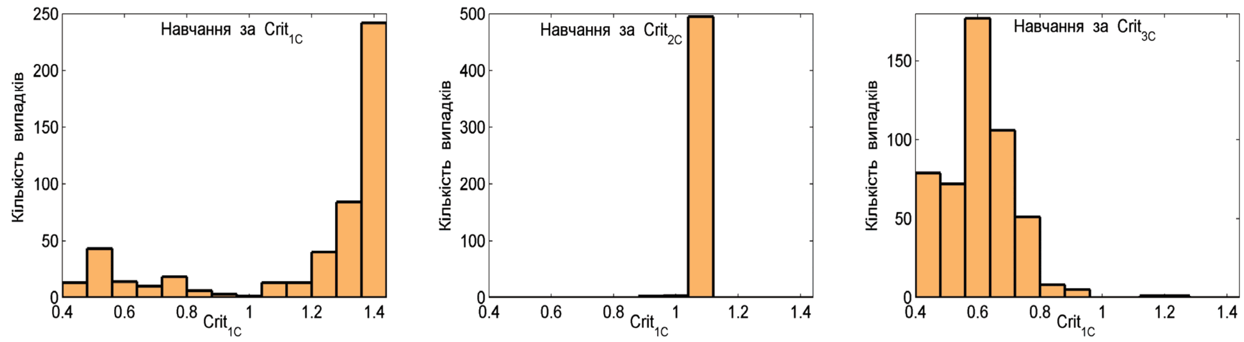

Рис. 1. Розподіл результатів навчання при настроюванні нечіткого класифікатора, *t*-норму якого реалізовано операцією мінімума

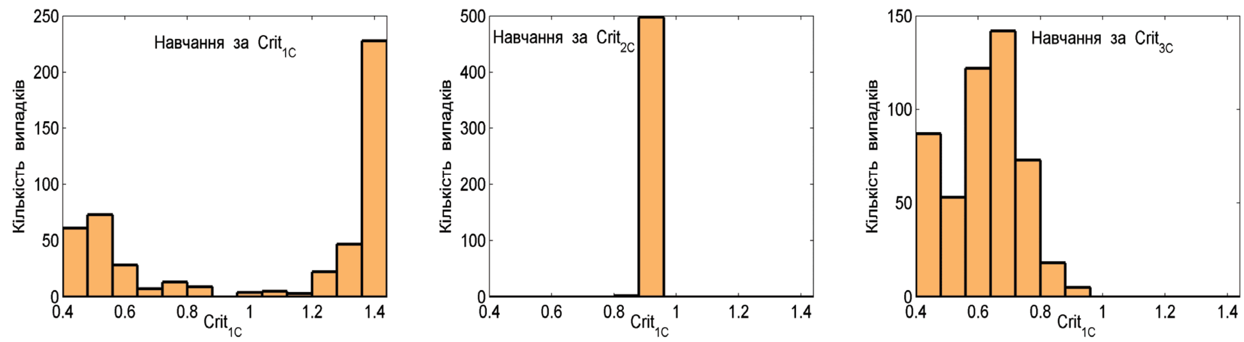

Рис. 2. Розподіл результатів навчання при настроюванні нечіткого класифікатора, *t*-норму якого реалізовано добутком

В експериментах з навчання за критерієм $Crit_{3C}$ штрафний коефіцієнт $(penalty)$ обирався випадково з діапазону $(0,10]$. Для визначення впливу його на якість навчання введемо коефіцієнт успішності α. Він показує частку процесів навчання, які закінчилися успіхом за фіксованого значення штрафного коефіцієнта. Успішним назвемо процес навчання, результат якого входить в 20 % кращих. Для класифікатора, у якого *t*-норму реалізовано операцією мінімуму, після 1500 експериментів отримано 85 різних результатів навчання. Поріг, для входження в 20 % кращих, склав 0,568, що відповідає значенню критерію $Crit_{1C}$ для класифікатора з 16-м рангом. Для класифікатора, у якого *t*-норма реалізовану добутком, після 1500 експериментів отримано 82 різних результатів навчання. Поріг, для входження в 20 % кращих, становить 0,56, що відповідає значенню критерію $Crit_{1C}$ для класифікатора з 16-м рангом. Розподіл результатів експериментів (рис. 3, 4) показує, що найчастіше успішне навчання відбувається коли $penalty \in (0,\ 2{,}5]$.

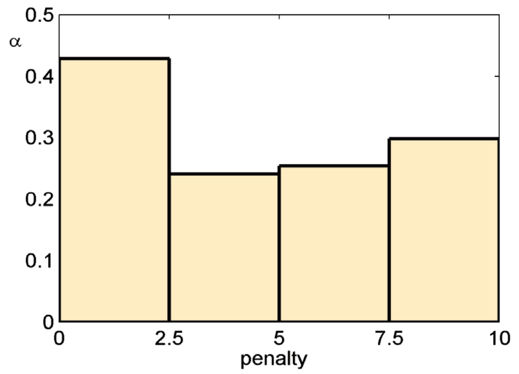

Рис. 3. Залежність успішності навчання від штрафного коефіцієнта для нечіткого класифікатора, *t*-норму якого реалізовано операцією мінімума

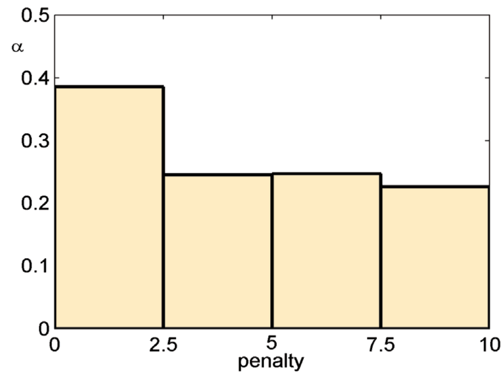

Рис. 4. Залежність успішності навчання від штрафного коефіцієнта для нечіткого класифікатора, *t*-норму якого реалізовано добутком

Висновки

Узагальнено критерії навчання нечіткого класифікатора на випадок платіжної матриці, в якій записані вартості помилок різних типів. Комп'ютерні експерименти із розв'язання задачі діагностики хвороби серця за даними з Statlog Heart Data Set показали, що найкращу якість настроювання забезпечує використання критерію навчання, в якому відстань між нечіткими результатами логічного виведення та експериментальними даними для випадків помилкової класифікації зважується штрафним коефіцієнтом.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Штовба С. Д. Порівняння критеріїв навчання нечіткого класифікатора / С. Д. Штовба // Вісник Вінницького політехнічного інституту. — 2007. — № 6. — С. 84—91.

2. Scott C. Neyman-Pearson approach to statistical learning / Scott C., Nowak R. A // IEEE Transactions on Information Theory. — 2005. — Vol. 51, № 11. — P. 3806—3819.

3. Штовба С. Д. Проектирование нечетких систем средствами MATLAB / С. Д. Штовба. — М. : Горячая линия — Телеком, 2007. — 288 с.

4. Ishibuchi H. Voting in fuzzy rule-based systems for pattern classification problems/ Ishibuchi H., Nakashima T., Morisawa T. // Fuzzy Sets and Systems. — 1999. — Vol. 103, № 2. — P. 223—238.

5. Rotshtein A. Design and tuning of fuzzy rule–based system for medical diagnosis. In «Fuzzy and Neuro-Fuzzy Systems in Medicine» (Eds. : Teodorescu N. H., Kandel A., and Jain L. C.) / A. Rotshtein. Boca-Raton : CRC-Press, 1998. — P. 243—289.

6. Shtovba S. Tuning the fuzzy classification models with various learning criteria: the case of credit data classification / Shtovba S., Pankevich O., Dounias G. // Proc. of Inter. Conference on Fuzzy Sets and Soft Computing in Economics and Finance. St. Petersburg (Russia), 17—20 June 2004. — Vol. 1. — St. Petersburg : Russian Fuzzy Systems Association, 2004. — P. 103—110.

7. Растригин Л. А. Адаптация сложных систем. Методы и приложения / Л. А. Растригин. — Рига : Зинатне, 1981. — 375 с.

8. Bache K., Lichman M. UCI Machine Learning Repository [Электронный ресурс] / Bache K., Lichman M. Irvine, CA: University of California, School of Information and Computer Science. — 2013. — Режим доступу : http://archive.ics.uci.edu/ml.

9. Штовба С. Д. Обеспечение точности и прозрачности нечеткой модели Мамдани при обучении по экспериментальным данным / С. Д. Штовба // Проблемы управления и информатики. — 2007. — № 4. — С. 102—114.

Рекомендована кафедрою комп'ютерних систем управління

Стаття надійшла до редакції 18.10.2013
Рекомендована до друку 28.10.2013

Штовба Сергій Дмитрович — професор, ***Нагорна Анастасія Володимирівна*** — аспірантка.
Кафедра комп'ютерних систем управління;
Панкевич Ольга Дмитрівна — доцент кафедри теплогазопостачання;
Вінницький національний технічний університет, Вінниця